Structure from Motion

Raul Queiroz Feitosa

Objectives

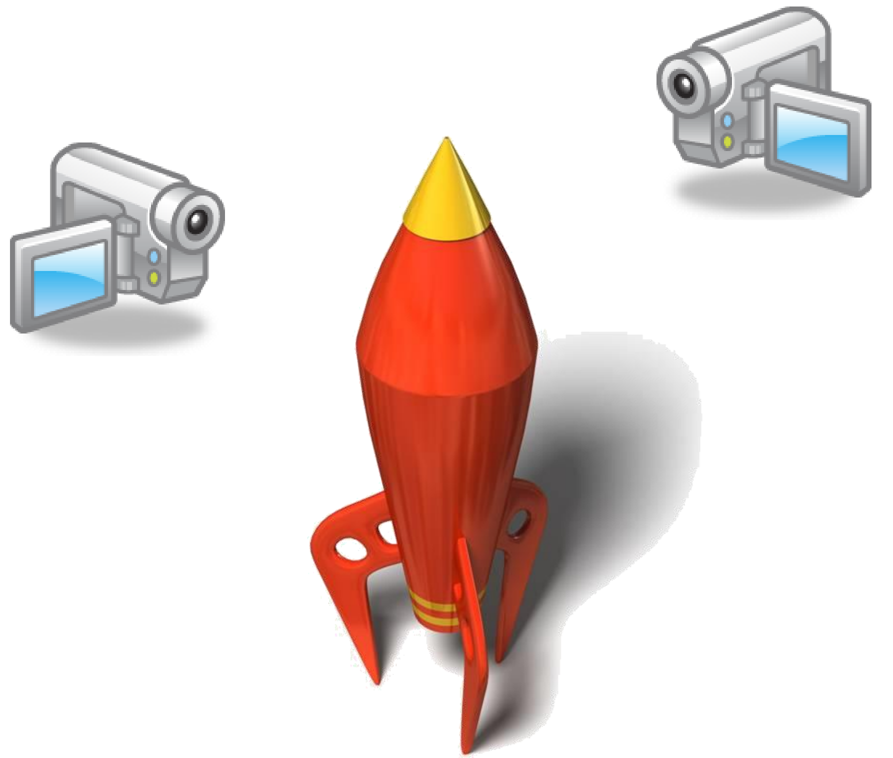
- *State the Structure from Motion (SFM) problem,*
- *Introduce SFM techniques for small relief (affine), and*
- *Introduce SFM techniques for large relief (projective).*

Contents

- Introduction
- Affine SFM
- Projective SFM

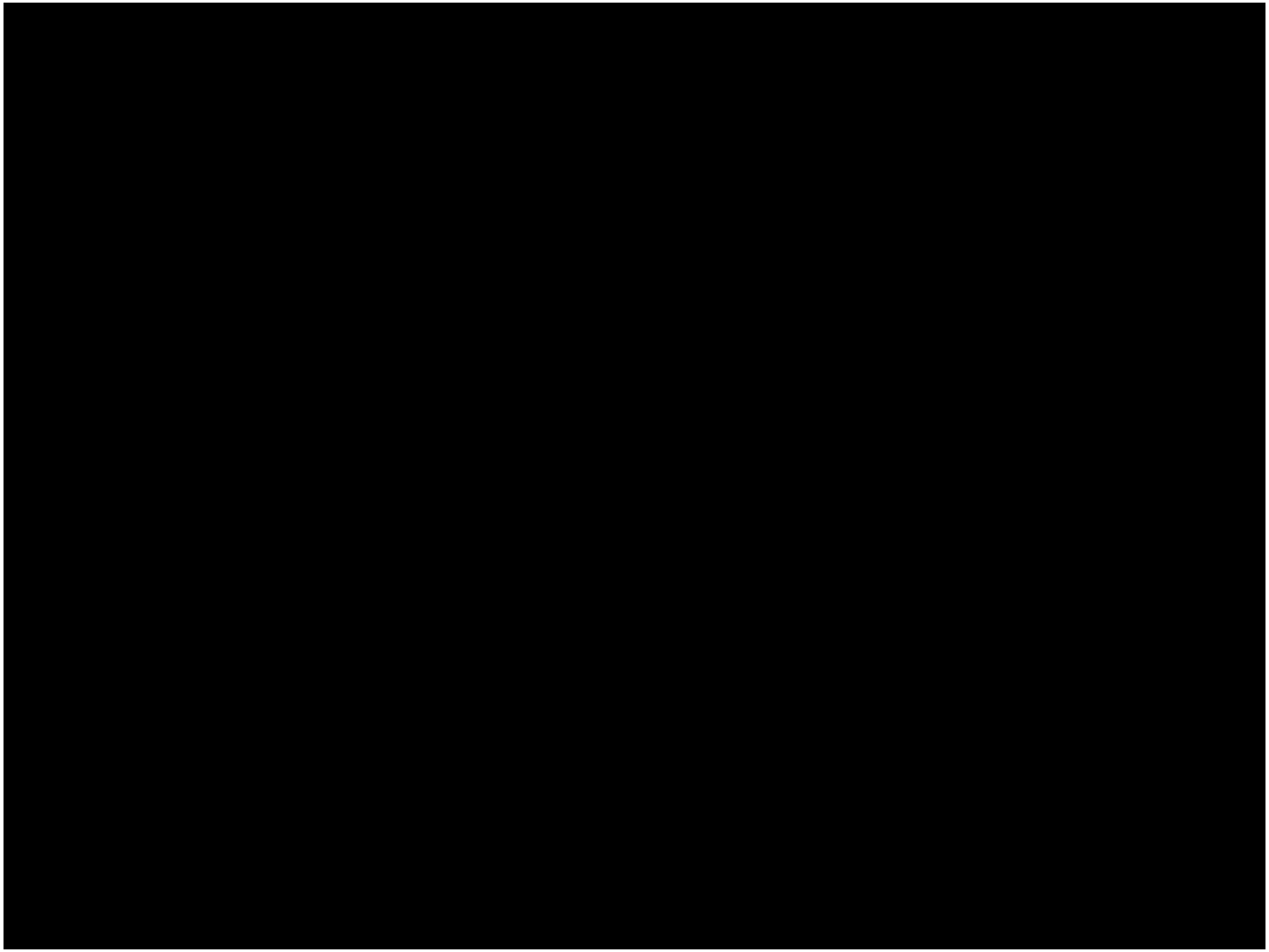
Structure from motion

Automatic recovery of *camera motion* and *scene structure* from two or more images.



also called automatic *camera tracking* or *matchmoving*.

Structure from motion



[*click here to watch the video*]

Applications of SFM

- Computer vision:
  - multiple-view shape reconstruction,
  - novel view synthesis,
  - autonomous vehicle navigation.
  - …
- Film production:
  - insertion of computer-generated imagery (CGI) into live-action backgrounds,
  - …

Stereopsis × SFM

- **Stereopsis**
  - Intrinsic camera parameters known (calibrated cameras), and
  - Extrinsic camera parameters known (camera pose relative to a fixed world coordinate system known)
- **Structure from Motion**
  - Camera parameters are (partially) unknown and
  - possibly change over time.

more difficult

SFM problem formulation

Given the images $\boldsymbol{p}_{ij}$ seen by m cameras of n fixed 3D points $\boldsymbol{P}_j$ (homogeneous)

estimate the m projection matrices $\mathcal{M}_i$ and the 3D points $\boldsymbol{P}_j$ from the mn correspondences $\boldsymbol{p}_{ij}$.

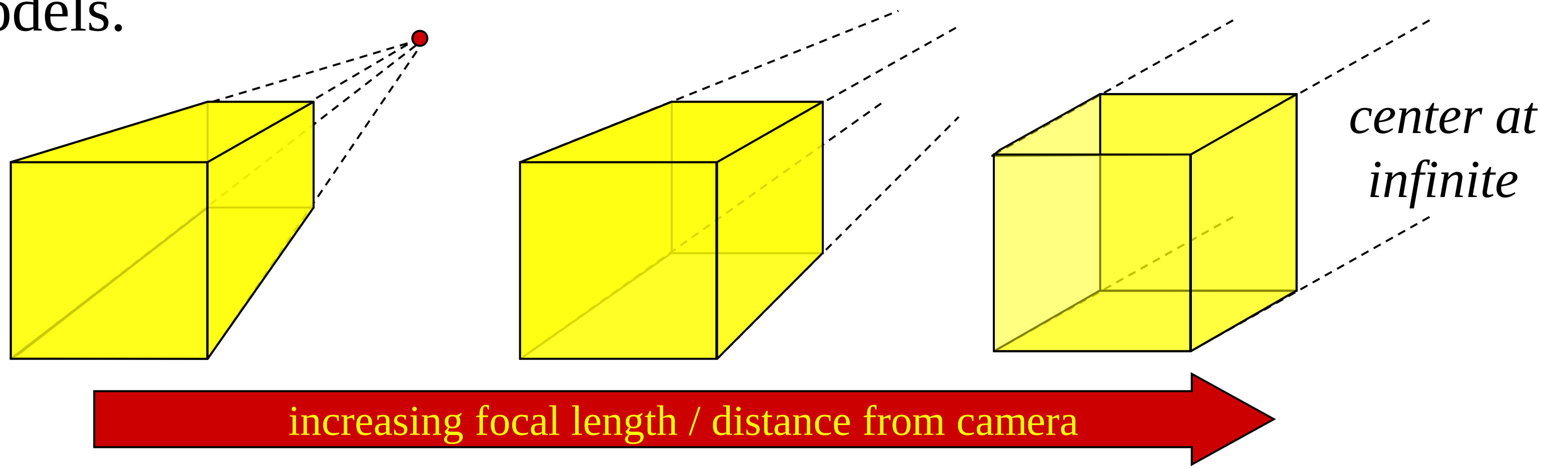
Structure from Motion

Contents:

- Introduction
- Affine SFM
- Projective SFM

Affine approximation

If scene's relief is small compared with distance z_j to the camera the perspective projection may be approximated by *affine* models.



Affine model

For affine cameras (*see chapter on camera models*), the previous equation takes the (nonhomogeneous) form

$$\boldsymbol{p}_{ij} = \mathcal{M}_i \begin{pmatrix} \boldsymbol{P}_j \\ 1 \end{pmatrix} = \mathcal{A}_i \boldsymbol{P}_j + \boldsymbol{b}_i$$

where $\mathcal{M}_i$ is the affine projection matrix, for $i=1, \ldots, m$, $j=1, \ldots, n$.

Dropping i and j from the notation above, yields

$$\boldsymbol{p} = \overset{2\times 1}{\begin{pmatrix} u \\ v \end{pmatrix}} = \overset{2\times 3}{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}} \overset{3\times 1}{\begin{pmatrix} x \\ y \\ z \end{pmatrix}} + \overset{2\times 1}{\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}} = \overset{2\times 3}{\mathcal{A}} \, \overset{3\times 1}{\boldsymbol{P}} + \overset{2\times 1}{\boldsymbol{b}}$$

Affine model

Thus

$$\overset{2\times 1}{\boldsymbol{p}_{ij}} = \overset{2\times 4}{\mathcal{M}_i} \begin{pmatrix} \overset{3\times 1}{\boldsymbol{P}_j} \\ 1 \end{pmatrix} = \overset{2\times 3}{\mathcal{A}_i} \overset{3\times 1}{\boldsymbol{P}_j} + \overset{2\times 1}{\boldsymbol{b}_i}$$

involves $2mn$ equations in $8m+3n$ unknowns.

Hence, for a sufficient number of views (m) and a sufficient number of points (n), *camera motion* ($\mathcal{M}_i$) and *scene structure* ($\mathbf{P}_i$) may be recovered.

Affine ambiguity

If $\mathcal{M}_i$ and $\boldsymbol{P}_j$ are solutions to the SFM affine problem, so are $\mathcal{M}_i'$ and $\boldsymbol{P}_j'$, where

$$\mathcal{M}_i' = \mathcal{M}_i Q \quad \text{and} \quad \begin{pmatrix} \boldsymbol{P}_j' \\ 1 \end{pmatrix} = Q^{-1} \begin{pmatrix} \boldsymbol{P}_j \\ 1 \end{pmatrix}$$

and $\mathcal{Q}$ is an arbitrary *affine transformation matrix*, - that is, it can be written as

$$Q = \begin{pmatrix} \mathcal{C} & \boldsymbol{d} \\ \mathbf{0}^T & 1 \end{pmatrix} \quad \text{with} \quad Q^{-1} = \begin{pmatrix} \mathcal{C}^{-1} & -\mathcal{C}^{-1}\boldsymbol{d} \\ \mathbf{0}^T & 1 \end{pmatrix},$$

where $\mathcal{C}$ is a nonsingular 3x3 matrix and $\boldsymbol{d} \in \mathbb{R}^3$.

Affine ambiguity

The affine SFM problem can *only be defined up to an affine transformation ambiguity*.

Taking into account the 12 parameters of the general affine transformation, a solution may exist as long as

$$\underbrace{2mn}_{} \geq 8\overset{\#\text{ of }\mathcal{M}_i}{m} + 3\overset{\#\text{ of }\boldsymbol{P}_j}{n} - 12$$

Thus, for 2 views, we need 4 point correspondences.

Offset to the center of mass

Translate the origin of the image coordinate system to the center of mass $\boldsymbol{p}_{i0}$ of the observed points ($\boldsymbol{p}_{ij} \rightarrow \boldsymbol{p}_{ij}$ - $\boldsymbol{p}_{i0}$), by doing

$$\boldsymbol{p}_{ij} \rightarrow \boldsymbol{p}_{ij} - \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{p}_{ik} = \mathcal{A}_i\boldsymbol{P}_j + \boldsymbol{b}_i - \frac{1}{n}\sum_{k=1}^{n}\left(\mathcal{A}_i\boldsymbol{P}_k + \boldsymbol{b}_i\right) = \mathcal{A}_i\left(\boldsymbol{P}_j - \frac{1}{n}\sum_{k=1}^{n}\boldsymbol{P}_k\right)$$

For simplicity, the origin of the world coordinate system is moved to the centroid of the 3D points ($\boldsymbol{P}_j \rightarrow \boldsymbol{P}_j$ - $\boldsymbol{P}_0$)

This yields

$$\boldsymbol{p}_{ij} = \mathcal{A}_i\boldsymbol{P}_j$$

The measurement matrixx

Let's create a $2m \times n$ data (measurement) matrix $\mathcal{D}$:

$$\mathcal{D} = \begin{array}{c c} & \text{points } (n) \longrightarrow \\ \text{cameras } (2m) \downarrow & \overbrace{\begin{pmatrix} \boldsymbol{p}_{11} & \boldsymbol{p}_{12} & \cdots & \boldsymbol{p}_{1n} \\ \boldsymbol{p}_{21} & \boldsymbol{p}_{22} & \cdots & \boldsymbol{p}_{2n} \\ & & \ddots & \\ \boldsymbol{p}_{m1} & \boldsymbol{p}_{m2} & \cdots & \boldsymbol{p}_{mn} \end{pmatrix}}^{2m \times n} \end{array} = \overbrace{\begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \vdots \\ \mathcal{A}_m \end{pmatrix}}^{2m \times 3} \overbrace{\begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 & \cdots & \boldsymbol{P}_n \end{pmatrix}}^{3 \times n} = \mathcal{A}\mathcal{P}$$

The measurement matrix $\mathcal{D} = \mathcal{A}\mathcal{P}$ must have rank 3!

Tomasi-Kanade factorization

Applying *SVD* to matrix $\mathcal{D}$

$$\mathcal{D}=\mathcal{U}\mathcal{W}\mathcal{V}^T$$

Since rank($\mathcal{D}$)=3 (noise free)

$$\mathcal{U}=\begin{bmatrix} \mathcal{U}_3 & \mathcal{U}_{n-3} \end{bmatrix} \qquad \mathcal{W}=\begin{bmatrix} \mathcal{W}_3 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathcal{V}^T=\begin{bmatrix} \mathcal{V}^T{}_3 \\ \mathcal{V}^T{}_{n-3} \end{bmatrix}$$

Therefore

$$\underset{2m\times n}{\mathcal{D}}=\underset{2m\times 3}{\mathcal{U}_3}\,\underset{3\times 3}{\mathcal{W}_3}\,\underset{3\times n}{\mathcal{V}^T{}_3}=\mathcal{A}\mathcal{P}$$

In practice, due to noise, inaccuracy in location of the interest points, and the mere fact that cameras are not affine, the matrix $\mathcal{D}$ is full rank.

Tomasi-Kanade factorization algorithm

1. Compute the singular value decomposition.

$$\mathcal{D}=\mathcal{U}\mathcal{W}\mathcal{V}^T$$

2. Construct the matrices $\mathcal{U}_3$, $\mathcal{V}_3$ and $\mathcal{W}_3$ formed by the three leftmost columns of the matrices $\mathcal{U}$ and $\mathcal{V}$ (the ones corresponding to the largest singular values), and the corresponding 3×3 sub matrix of $\mathcal{W}$.

3. Define

$$\mathcal{A}_0 = \mathcal{U}_3\ (\mathcal{W}_3)^{½} \qquad\qquad \mathcal{P}_0 = (\mathcal{W}_3)^{½}\ \mathcal{V}^T{}_3$$

the $2m$×3 matrix is an estimate of the camera motion, and the 3×n matrix is an estimate of the scene structure.

It can be proven that $\mathcal{W}_3$ is the closest rank-3 approximation of $\mathcal{D}$.

Recalling the affine ambiguity

The decomposition

$$\mathcal{A}_o = \mathcal{U}_3 \, (\mathcal{W}_3)^{½} \qquad \mathcal{P}_o = (\mathcal{W}_3)^{½} \, \mathcal{V}^T{}_3$$

is not unique.

For any 3×3 non singular matrix $\mathcal{C}$

$$\mathcal{A} = \mathcal{A}_o \, \mathcal{C} \quad \text{and} \quad \mathcal{P} = \mathcal{C}^{-1} \, \mathcal{P}_o$$

is also a solution (affine ambiguity).

Affine camera models

Summary of **affine mappings**

$$\boldsymbol{p} = \mathcal{M}\,{}^{W}\boldsymbol{P}\,, \quad \text{where}$$

$$\mathcal{M} = \begin{pmatrix} \boldsymbol{r}_1^T & t_x \\ \boldsymbol{r}_2^T & t_y \\ \mathbf{0} & 1 \end{pmatrix}$$

***orthographic** - 5 dof*

$$\mathcal{M} = \alpha \begin{pmatrix} \boldsymbol{r}_1^T & t_x \\ \boldsymbol{r}_2^T & t_y \\ \mathbf{0} & 1/\alpha \end{pmatrix}$$

***scaled orthographic** – 6 dof*

$$\mathcal{M} = \begin{pmatrix} \alpha\boldsymbol{r}_1^T & \alpha t_x \\ \beta\boldsymbol{r}_2^T & \beta t_y \\ \mathbf{0} & 1 \end{pmatrix}$$

***weak-perspective** - 7 dof*

$$\mathcal{M} = \begin{pmatrix} \alpha & -\alpha\cot\theta & 0 \\ & \beta/\sin\theta & 0 \\ & & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{r}_1^T & t_x \\ \boldsymbol{r}_2^T & t_y \\ \mathbf{0} & 1 \end{pmatrix}$$

***general affine** - 8 dof*

Note that in all cases the 3rd row is =(0 0 0 1)!

Eliminating the affine ambiguity

From a previous slide

$$\boldsymbol{p}_{ij} = \mathcal{A}_i \boldsymbol{P}_j$$

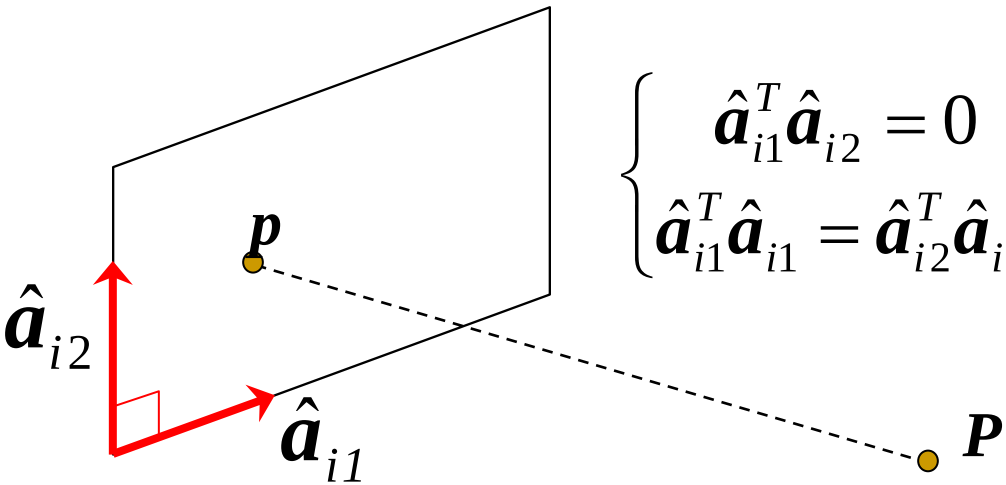
the image coordinates $\boldsymbol{p}_{ij}$ are the projections of $\boldsymbol{P}_{ij}$, on the vectors $\boldsymbol{a}^T{}_{i1}$ and $\boldsymbol{a}^T{}_{i2}$, the rows of $\mathcal{A}_i$.

Consider scaled orthographic projection: image axes are perpendicular.

$$\begin{cases} \hat{\boldsymbol{a}}_{i1}^T \hat{\boldsymbol{a}}_{i2} = 0 \\ \hat{\boldsymbol{a}}_{i1}^T \hat{\boldsymbol{a}}_{i1} = \hat{\boldsymbol{a}}_{i2}^T \hat{\boldsymbol{a}}_{i2} \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} = 0 \\ \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i1} = \boldsymbol{a}_{i2}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} \end{cases}$$

Eliminating the affine ambiguity

Scaled orthographic: image axes are perpendicular



$$\begin{cases} \hat{\boldsymbol{a}}_{i1}^T \hat{\boldsymbol{a}}_{i2} = 0 \\ \hat{\boldsymbol{a}}_{i1}^T \hat{\boldsymbol{a}}_{i1} = \hat{\boldsymbol{a}}_{i2}^T \hat{\boldsymbol{a}}_{i2} \end{cases} \Leftrightarrow \begin{cases} \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} = 0 \\ \boldsymbol{a}_{i1}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i1} = \boldsymbol{a}_{i2}^T \mathcal{C}\mathcal{C}^T \boldsymbol{a}_{i2} \end{cases}$$

This translates into $2m$ quadratic equations in the elements of $\mathcal{C}$

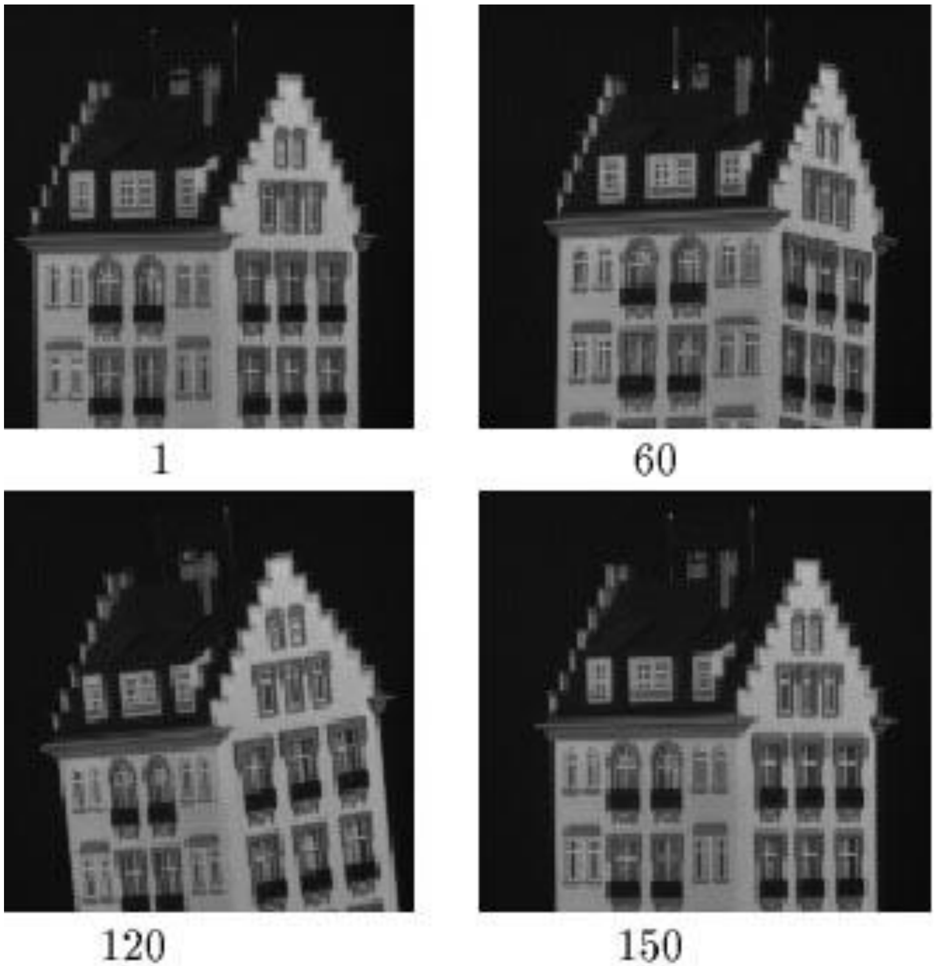
- Solve for $\mathcal{C}$
- Update $\mathcal{A}$ and $\mathcal{P} \rightarrow$ $\mathcal{A} = \mathcal{A}_o\mathcal{C}$ and $\mathcal{P} = \mathcal{C}^{-1}\mathcal{P}_o$
- Note that the solution is defined up to an arbitrary rotation.
- Common practice: $\rightarrow$ first camera frame $\equiv$ world frame.

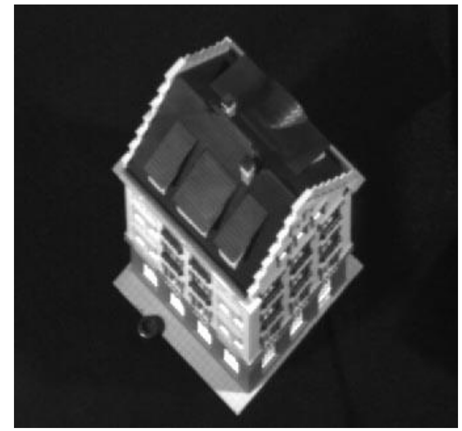
Tomasi-Kanade algorithm summary

- Given: m images and n features $\boldsymbol{p}_{ij}$
- For each image i, center the feature coordinates
- Construct a $2m \times n$ measurement matrix $\mathcal{D}$:
  - Column j has the projection of point $\boldsymbol{P}_j$ in all views
  - Row i has one coordinate of the projections of all n points in image i
- Factorize $\mathcal{D}$:
  - Compute SVD: $\mathcal{D} = \mathcal{U}\ \mathcal{W}\ \mathcal{V}^T$
  - Create $\mathcal{U}_3$ by taking the first 3 columns of $\mathcal{U}$
  - Create $\mathcal{V}_3$ by taking the first 3 columns of $\mathcal{V}$
  - Create $\mathcal{W}_3$ by taking the upper left 3 × 3 block of $\mathcal{W}$
- Create the motion and shape matrices:
  - $\mathcal{A}_o = \mathcal{U}_3\mathcal{W}_3^{1/2}$ and $\mathcal{P}_o = \mathcal{W}_3^{1/2}\ \mathcal{V}_3^{T}$ (or $\mathcal{A}_o = \mathcal{U}_3$ and $\mathcal{P}_o = \mathcal{W}_3\mathcal{V}_3^{T}$)
- Eliminate affine ambiguity.

Source: M. Hebert

Reconstruction results





Figures from
C. Tomasi and T. Kanade.
[Shape and motion from image streams under orthography: A factorization method.]
International Journal of Computer Vision, 9(2):137-154, November 1992.

Contents

- Introduction
- Affine SFM
- Projective SFM

Projective model

For projective cameras we have

$$\boldsymbol{p}_{ij} = \begin{pmatrix} u_{ij} \\ v_{ij} \\ 1 \end{pmatrix} = \frac{1}{z_{ij}} \mathcal{M}_i \boldsymbol{P}_j \quad \text{or} \quad \begin{cases} u_{ij} = \dfrac{\boldsymbol{m}_{i1}\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\boldsymbol{P}_j} \\ v_{ij} = \dfrac{\boldsymbol{m}_{i2}\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\boldsymbol{P}_j} \end{cases}$$

for $i=1,\ldots,m$ and $j=1,\ldots,n$ where $\boldsymbol{m}^T{}_{i1}$, $\boldsymbol{m}^T{}_{i2}$, $\boldsymbol{m}^T{}_{i3}$, are the rows of a 3×4 matrix $\mathcal{M}_i$ and $\boldsymbol{P}_j$ denotes the homogeneous coordinate vector of the point P_j in some world coordinate system.

Projective ambiguity

$$\boldsymbol{p}_{ij} = \begin{pmatrix} u_{ij} \\ v_{ij} \\ 1 \end{pmatrix} = \frac{1}{z_{ij}} \mathcal{M}_i \boldsymbol{P}_j \quad \text{or} \quad \begin{cases} u_{ij} = \dfrac{\boldsymbol{m}_{i1}\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\boldsymbol{P}_j} \\ v_{ij} = \dfrac{\boldsymbol{m}_{i2}\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\boldsymbol{P}_j} \end{cases}$$

According to Theorem 1 (*see chapter on Camera Model*) $\mathcal{M}_i$ is an arbitrary rank-3 3×4 matrix.

Hence, if $\mathcal{M}_i$ and $\boldsymbol{P}_j$ are solutions for the equation above, so are

$$\mathcal{M}'_i = \mathcal{M}_i \mathcal{Q} \quad \text{and} \quad \boldsymbol{P}'_j = \mathcal{Q}^{-1} \boldsymbol{P}_j ,$$

where $\mathcal{Q}$ is an arbitrary *projective transformation matrix*, i.e., any non singular 4×4 matrix.

Projective ambiguity

The 4×4 matrix $\mathcal{Q}$ is defined up to a scale, since multiplying it by a nonzero scalar amounts to applying inverse scaling to $\mathcal{M}_i$ and $\boldsymbol{P}_j$.

Thus, the projective SFM problem can *only be defined up to a projective transformation ambiguity*.

Taking into account the *projective ambiguity* the problem has a finite number of solutions as soon as

$$2mn \geq \underbrace{11m}_{\#\text{ of } \mathcal{M}_i}+\underbrace{3n}_{\#\text{ of } \boldsymbol{P}_j}-15$$

For two views, we need seven point correspondences.

Bilinear projection SFM

As in algebraic reconstruction, we do

$$z_{ij}\,\boldsymbol{p}_{ij} = \mathcal{M}_i\,\boldsymbol{P}_j \Rightarrow \quad \boldsymbol{p}_{ij} \times \mathcal{M}_i\,\boldsymbol{P}_j = \mathbf{0}$$

So, we look for a solution that minimizes

$$E = \sum_{ij} \left\| \boldsymbol{p}_{ij} \times \mathcal{M}_i \boldsymbol{P}_j \right\|^2$$

This can be done by alternating steps where $\boldsymbol{P}_j$ are kept constant (estimated) while $\mathcal{M}_i$ are estimated (kept constant).

Bundle Adjustment

Given estimates for the matrices $\mathcal{M}_i$ ($i=1,\ldots,m$) and vectors $\boldsymbol{P}_j$ ($j=1,\ldots,n$) , refine them by using non linear least square to minimize the global reprojection error E:

$$E = \frac{1}{mn}\sum_{i,j}\left[\left(u_{ij} - \frac{\boldsymbol{m}_{i1}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j}\right)^2 + \left(v_{ij} - \frac{\boldsymbol{m}_{i2}\cdot\boldsymbol{P}_j}{\boldsymbol{m}_{i3}\cdot\boldsymbol{P}_j}\right)^2\right]$$

Expensive, but encompasses a physically significant error measure.

As initial guesses use $\boldsymbol{P}_j$ from Affine SFM e and from them compute $\mathcal{M}_i$.

Eliminating the projective ambiguity

Let us assume that $\mathcal{M}_i$ ($i=1,\ldots,m$) and vectors $\boldsymbol{P}_j$ ($j=1,\ldots,n$) have been estimated.

If $\hat{\mathcal{M}}_i$ and $\hat{\boldsymbol{P}}_j$ are a reconstruction in a particular *Euclidean coordinate system*, there exist a 4×4 matrix $\mathcal{Q}$ such that

$$\hat{\mathcal{M}}_i = \mathcal{M}_i \mathcal{Q} \quad \text{and} \quad \hat{\boldsymbol{P}}_i = \mathcal{Q}^{-1} \boldsymbol{P}_i$$

The next slides show two methods of computing the *Euclidean upgrade matrix* $\mathcal{Q}$, when (some of) the intrinsic parameters are known.

Eliminating the projective ambiguity

Method 1 (calibrated cameras):

- $\hat{\mathcal{M}}_i$ is defined up to a scale, so

$$\hat{\mathcal{M}}_i = \rho_i \mathcal{K}_i (\mathcal{R}_i \;\; \boldsymbol{t}_i)$$

  where ρ_i accounts for the scale and $\mathcal{K}_i$ is the calibration matrix.

- Writing $\mathcal{Q}=(\mathcal{Q}_3\ \boldsymbol{q}_4)$, where $\mathcal{Q}_3$ is 4×3 matrix and $\boldsymbol{q}_4$ is a vector in $\mathbb{R}^4$, yields

$$\mathcal{M}_i \mathcal{Q}_3 = \rho_i \mathcal{K}_i \mathcal{R}_i$$

- If $\mathcal{K}_i$ is known (calibrated), $\mathcal{K}_i^{-1} \mathcal{M}_i \mathcal{Q}_3 = \rho_i \mathcal{R}_i$ can be treated as a scaled rotation matrix, Thus ….

Eliminating the projective ambiguity

Method 1 (cont.):

- … the rows $\boldsymbol{m}_{ij}^T$ ($j=1,2,3$) of $\mathcal{K}_i^{-1}\mathcal{M}_i$ are perpendicular vectors, so

$$\begin{cases} \boldsymbol{m}_{i1}^T Q_3 Q_3^T \boldsymbol{m}_{i2} = 0 \\ \boldsymbol{m}_{i2}^T Q_3 Q_3^T \boldsymbol{m}_{i3} = 0 \\ \boldsymbol{m}_{i3}^T Q_3 Q_3^T \boldsymbol{m}_{i1} = 0 \\ \boldsymbol{m}_{i1}^T Q_3 Q_3^T \boldsymbol{m}_{i1} - \boldsymbol{m}_{i2}^T Q_3 Q_3^T \boldsymbol{m}_{i2} = 0 \\ \boldsymbol{m}_{i2}^T Q_3 Q_3^T \boldsymbol{m}_{i2} - \boldsymbol{m}_{i3}^T Q_3 Q_3^T \boldsymbol{m}_{i3} = 0 \end{cases}$$

where $\mathcal{Q}_3$ is defined up to a scale. To determine it uniquely we assume that the world coordinate system coincides with the first camera's frame, that means, $\mathcal{K}_i^{-1}\mathcal{M}_i$=[Id **0**].

•

Eliminating the projective ambiguity

Method 1 (cont.):

- Given m images, we obtain $5(m-1)$ quadratic equations in the coefficients of $\mathcal{Q}$ that can be solved using non linear least squares.
- The vector $\boldsymbol{q}_4$ of matrix $\mathcal{Q}$

$$\mathcal{Q}=(\mathcal{Q}_3\ \boldsymbol{q}_4),$$

can be determined by assuming (arbitrarily) that the origin of world coordinate system and the origin of the first camera's frame coincide.

Eliminating the projective ambiguity

Method 2:

- The previous equations are linear in the **10** coefficients of the symmetric matrix $\mathcal{A}=\mathcal{Q}_3\mathcal{Q}_3^T \rightarrow$ (linear least squares)
- To enforce the rank 3 of $\mathcal{A}$, we do

$$\mathcal{A}=\mathcal{U}\Lambda\mathcal{U}^T,$$

where Λ is the diagonal matrix with the eigenvalues of $\mathcal{A}$ and $\mathcal{U}$ is the orthogonal matrix formed by its eigenvectors.

Eliminating the projective ambiguity

Method 2 (cont.):

- To enforce rank 3 of $\mathcal{A}$ we do

$$\mathcal{Q}=\mathcal{U}_3\Lambda_3^{1/2}$$

  where $\mathcal{U}_3$ is formed by the 3 columns of $\mathcal{U}$ associated to the 3 largest eigenvalues of $\mathcal{A}$ and Λ_3 is the corresponding sub matrix of Λ.

- Vector $\boldsymbol{q}_4$ can be computed as in the previous method.

Eliminating the projective ambiguity

These methods can be adapted to the case where only some of the intrinsic camera parameters are known.

Since $\mathcal{R}_i$ are orthogonal matrices

$$\mathcal{M}_i \mathcal{A} \mathcal{M}_i^T = \rho_i^2 \mathcal{K}_i \mathcal{K}_i^T$$

each image provides a set of constraints between the entries of $\mathcal{R}_i$ and $\mathcal{A}$.

Eliminating the projective ambiguity

Assuming, for example that the center of the image is known for each camera, we can take $u_0=v_0=0$ and

$$\mathcal{K}_i\mathcal{K}_i^T = \begin{pmatrix} \alpha_i^2 \dfrac{1}{\sin^2\theta_i} & -\alpha_i\beta_i \dfrac{\cos^2\theta_i}{\sin^2\theta_i} & 0 \\ -\alpha_i\beta_i \dfrac{\cos^2\theta_i}{\sin^2\theta_i} & \beta_i^2 \dfrac{1}{\sin^2\theta_i} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Each zero entry of $\mathcal{K}_i\mathcal{K}_i^T$ provides 2 independent linear equations in the 10 coefficients of the 4×4 matrix $\mathcal{A}$. With $m \geq 5$ images, these parameters can be estimated via linear least squares.

$$\begin{cases} \boldsymbol{m}_{i1}^T \mathcal{A} \boldsymbol{m}_{i3} = 0 \\ \boldsymbol{m}_{i2}^T \mathcal{A} \boldsymbol{m}_{i3} = 0 \end{cases}$$

What if zero skew can be assumed?

Readings

- C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137-154, November 1992.
- F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. Segmenting, Modeling, and Matching Video Clips Containing Multiple Moving Objects. PAMI 2007.
- Muhamad, S. and Hebert, M. (2000), Iterative Projective Reconstruction from Multiple Views, *Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. II, 430-437*.
- *Computer Vision - A modern approach* 2nd ED , by D. Forsyth and J. Ponce, 2012, chapter 8.

Next Topic

Kalman Filter